

# Medical Insurance Claims as a Source of Data for Research: Accuracy of Diagnostic Coding

Robert M. Worth MD, PhD\*, Robert E. Mytinger DrPH\*\*

*Validation of diagnostic codes in a sample of Hawaii medical insurance claims in 1986 to 1987 revealed 96% accuracy in hospital claims, high enough to supply data for research purposes. In physician claims the accuracy was only 62%. Initiation of two feedback loops to physicians from the insurer in 1989 resulted in a marked improvement of diagnostic coding accuracy by 1992 to 1994.*

## Introduction

Because of the high cost of generating original data, health service researchers often seek reliable secondary data for their studies. One appealing source of secondary data is claims submitted by hospitals and physicians to health insurance plans. Such claims typically contain encounter-specific information pertaining to procedures and other services provided to patients, to charges for such services, and to primary and secondary diagnoses. These variables, coupled with others descriptive of the patient and provider, form a potentially valuable resource with which to examine many questions pertinent to the delivery of health services. Questions of claims data accuracy have become important locally because of a major claims-based research data base developed to support health services research in Hawaii.

## Pretest

During 1987 planning began for establishing a continuing archive of abstracts of medical insurance claims processed by the Hawaii Medical Service Association (HMSA). After receiving the approval of the Hawaii Medical Association, a cooperative effort was begun by HMSA, the Hawaii Department of Health (DOH), the Department of Human Services (DHS), and the U.S. Centers for Disease

Control (CDC). The archival file, which is maintained by the DOH, has accrued data weekly from HMSA since late 1989. This data base is rigorously controlled, with access limited to qualified investigators seeking specific data elements for studies of substantial merit. To assure complete confidentiality, it contains no data by which the identity of individual patients or providers can be discerned by the researcher.

Of special concern has been the accuracy of the diagnostic data contained in claims submitted, so a pretest was done to validate the accuracy of the primary diagnostic codes being submitted by hospitals and physicians. This began by identifying medical insurance claims processed at HMSA between July 1986 and June 1987 which contained primary diagnostic codes (ICD9-CM) that matched a CDC-derived list of codes thought to be of surveillance interest. The 169,586 claims thus identified served as a sampling frame for a random sample (stratified by patient age and gender) of 186 claims for validation of the primary diagnostic code. With approval of the Hawaii Medical Association, an explanation of the project and a coding sheet were provided in the fall of 1987 to each hospital and physician's office that had originally submitted a sampled claim, together with a request for verification of the diagnosis on that hospital or physician claim.

Response to these requests was excellent. Hospital claims (N=47) reflected a very high level of accuracy, with only two hospital claims (4%) revealing coding errors, both typographical (Fig 1). While the blind re-coding of 38% of hospital claims did not reflect precisely the original codes, these represented *unprecise codes*, due almost exclusively to selection at the time of re-coding of another closely related condition as the *primary diagnosis*. These unprecise codes appeared to be largely for patients who had been admitted for several interrelated chronic problems. It is important to note that this validation study was performed on claims submitted prior to the mandated use of diagnostic-related group (DRG) hospital coding, which tends to minimize this problem. The remaining 57% of the re-coded hospital claims were identical with the diagnostic code originally submitted to HMSA. Hence, hospital-reported diagnoses appeared to be quite dependable for research purposes, probably reflecting a high level of training of hospital record room staff for diagnostic coding.

Physician-generated claims revealed a different picture, probably because the clerical staff in physician offices or billing service bureaus is less well-trained. Twelve percent of office claims appeared to have single digit typographical errors. Another 8% represented confusion of medical terminology. For example, a patient with *multiple myeloma* coded on the claim was changed to *malignant melanoma* on validation. Another class of errors in 17% of the office claims arose from coding a potential infectious disease instead of the exposure to that disease. For example, validation

\*Hawaii MEDTEP Research Center  
Pacific Health Research Institute

\*\*HMSA Foundation

Reprint requests:  
Dr Robert M. Worth  
Hawaii MEDTEP Research Center  
Pacific Health Research Institute  
846 South Hotel Street, Suite 302  
Honolulu, Hawaii 96813

A project supported in part by George F. Straub Trust, Hawaii Community Foundation, through the Pacific Health Research Institute; also supported by the HMSA Foundation.

Fig 1.—Diagnostic Coding Replicability of Insurance Claim Forms Versus Blind Re-coding of Medical Record Abstracts, HMSA, 1986 to 1987.

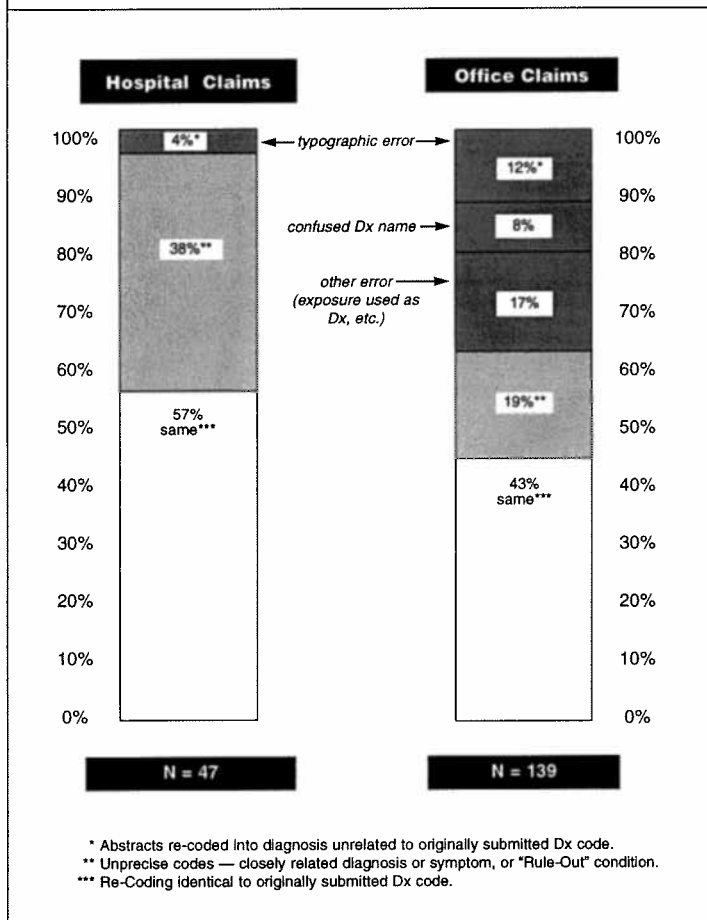
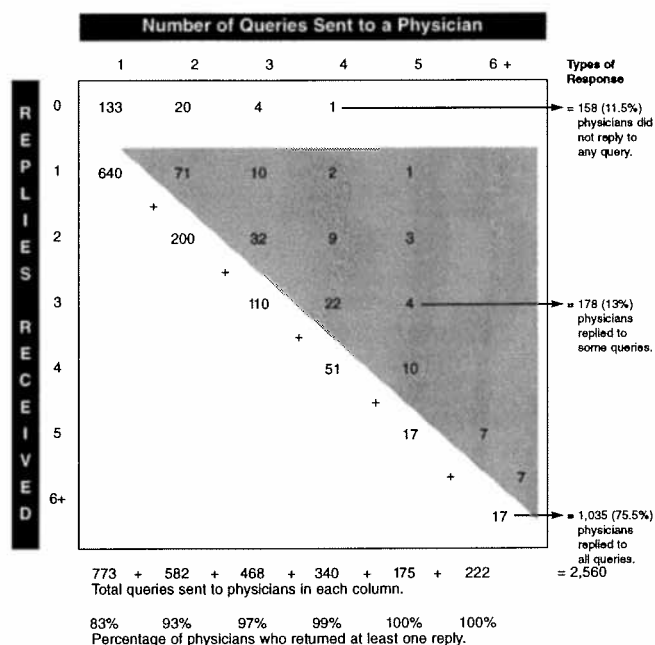


Table 1.—Distribution of 1,371 Hawaii physicians whose HMSA medical insurance claims were randomly queried during January 1992 to July 1994 for verification of the principal diagnostic code by number of replies received.\*



revealed that *dog-bite of the hand* had originally been coded as *rabies* (in a rabies-free state!). The selection of another related diagnosis from among several problems in the patient (eg, the *unprecise coding* category) was present in 19% of the office claims sampled. Thus, only 43% of the physician office claims were validated as having codes identical to the original code.

Since physician office claims represent the lion's share of the planned archival data base, a diagnostic accuracy rate (even stretching the point) of 62% made their utility for research purposes clearly questionable.

## Methods — Interventions

It was, therefore, decided to introduce two actions, each of which might improve the accuracy of diagnostic coding. The first was to broaden the system already in use by all claims processors, namely to query claims with improbable diagnoses. For example, age/gender algorithms already in place will flag a claim for a male with a diagnosis coded as *endometriosis*. A claim thus flagged is queried, with payment deferred until an explanation or a correction is received. Because of frequent coding problems found with infectious diseases in the preliminary study, the DOH provided HMSA with a list of infectious diseases very unlikely to be treated in Hawaii (cholera, poliomyelitis, rabies, African sleeping sickness, etc). These were then added to the flagging algorithms already used at HMSA.

A second action to track coding errors caused by flaws in staff training was instituted in 1991. This entailed the selection of a small random sample (about 20 queries a week) of medical office claims that had cleared the routine claims processing algorithms (see above). A form letter was sent from HMSA each week to each of the 20 sampled physicians. They were asked for verification of the accuracy of the diagnosis found on the specified claim recently submitted to HMSA. The accompanying worksheet contained the patient's name and birthdate, date of service, the ICD code of the submitted primary diagnosis, and a brief text of that code. The physician was asked either to confirm via the supplied self-addressed, postage-paid envelope that the diagnosis on the claim was correct, or to enter the correct diagnosis. The intent of this measure was to bring to the attention of physicians that elements on the claim are being monitored for accuracy. It was hypothesized that upon discovering coding errors, physicians will take steps to prevent their repetition by the office or service bureau staff responsible.

## Results — Infectious disease flagging

A test run with the expanded infectious disease flagging system in early 1989 immediately identified several coding errors. The most spectacular was a pseudo-outbreak of about 20 cases of anthrax judging from claims coded 022 emanating from the office of one obstetrician. The last verified human or animal case of anthrax in Hawaii was reported shortly before World War I! When queried by

HMSA, it was discovered that the billing clerk in that office was unaware of the difference between 022 and the code V22, which indicates a normal pregnancy. The offer by HMSA to assist in additional training of that clerk was happily accepted, and the anthrax epidemic quickly subsided. Other clusters of coding errors, partly from lack of training and partly from typographical errors, were discovered in this fashion and were amicably resolved.

## Results — Random queries

Table 1 reveals the distribution of random queries sent to a sample of 1,371 Hawaii physicians between January 1992 and July 1994. The 1,213 physicians who responded to some or all of the queries include a majority of physicians in active fee-for-service private practice in Hawaii. This response rate was unusually robust, with only 158 (11.5%) of the queried physicians failing to reply to any coding query sent. Most of the non-respondents had received only one or two queries during the 31-month study period. Another 178 physicians (13%) responded inconsistently, but the receipt by a physician of five or more queries invariably elicited at least one reply. The vast majority (84%) of these *inconsistent* responders failed to reply to only one of the queries sent, while responding to the rest. The remaining 1,035 physicians (75.5%) replied to all queries received, a truly remarkable level of cooperation.

Table 2 displays a response rate ranging between 79% and 86% from initial queries sent to physicians who had not previously been randomly selected, showing no trend over time. The number of physicians still available to receive their first query is steadily decreasing as time goes by. The response rate (84%) for those who received two or more queries was in the same range as for those who received their first query.

Compared to the low levels of coding accuracy previously found in claims from physicians (Fig 1), the results shown in Table 3 suggest a large improvement, with 94% now being validated as correct. The unprecise coding category is now down to about 6%, leaving less than 1% with a typographical or other obvious coding error.

## Discussion

The remarkable improvement in medical office diagnostic coding accuracy helps to validate the utility of medical claims information for research purposes in Hawaii. The reasons for the great improvement are less clear. What is apparent is the simple act of calling coding errors to the attention of physicians, whether by claims flagging and subsequent inquiry, or by routine random sampling and verification, seems to have the desired effect of heightening attention to accuracy in claims.

Of importance to investigators who use diagnostic data from insurance claims is our experience with unprecise codes. Many claims reflect medical services provided to persons with chronic conditions. This usually means that the attending physician is dealing with multiple co-morbidities. At one encounter, the physician may properly report diagnosis A as primary, and for the same patient on a subsequent encounter may report diagnosis B (or C or D) as primary. And each could be accurate. Thus, investigators working with diagnostic data from insurance claims files must view *all* the diagnoses found in a sequence of claims in order to see the context of the disease or condition being studied. Only then can they properly accommodate shifting diagnoses for the same patient.

## Acknowledgments

We owe a debt of gratitude to the early support of the Centers for Disease Control and Prevention, particularly Drs Steven Blount and

Ben Truman. We wish to express our gratitude for the ongoing assistance provided by Yeiso Arakaki of HMSA and Jim Cooper of the DOH in carrying out these trials.

We especially are grateful to the 1,213 Hawaii physicians who responded to the queries during the past three years, without whom all would be for naught.

Table 2.—Response rates to 2,560 coding queries for verification of the Principal Diagnostic Code by physicians in Hawaii during January 1992 to July 1994, by year of request and by initial query versus subsequent queries.

	Year of the Initial Query to a Physician				All Subsequent Queries
	1992 Queries	1993	1994 (Jan-July)	All Init	
Number of Queries Sent	624 (52/mo.)	491 (41/mo.)	256 (37/mo.)	1,371	1,189
Responses	537	389	212	1,138	997
Response Rates	86%	79%	83%	83%	84%

Table 3.—Distribution of 2,135 responses to queries for verification of the primary diagnostic code on randomly selected medical insurance claims processed by HMSA during 1992 to July 1994.

Categories of Response	Number in That Category	Illustrations of Coding Changes
New Dx code with the 1st 3 digits intact, or checked as 'OK as-is' on the query letter.	2,004 (94%)	None
Shift to specify location, detail, or alternative related code for the same problem.	78 (3.7%)	789.0 → 536.8 Abdomin. → Dyspepsia symptoms
Shift to new code, apparently to describe another problem in the same patient.	50 (2.3%)	461.1 → 473.0 Acute → Chronic sinusitis
		414.9 → 427.3 Ischemic heart dis. → Atrial fibrilla.
		600 → 535.5 Benign prostatic hypertrophy → Gastritis
'Off-the-wall' change, apparently due to single digit typographical error.	3	* 216.9 → 716.9 Benign lesion of the skin → Arthropathy
Total Responses	2,135	* Probable typographical error.